Enabling Memory Reliability, Availability, and Serviceability Features on Dell PowerEdge Servers

The memory subsystems on Dell™ PowerEdge™ 1850, PowerEdge 2800, and PowerEdge 2850 servers are designed to support reliability, availability, and serviceability (RAS) features such as error-correcting code, chip fail correct, spare banks, and mirroring. This article describes RAS features in detail, explaining how they are enabled, how they can affect available system memory, and how they can help to minimize system downtime caused by memory errors.

BY QINGSONG LI AND UTPAL PATEL

Related Categories:

Dell PowerEdge servers

Memory

System architecture

Visit www.dell.com/powersolutions for the complete category index.

Reliability, availability, and serviceability (RAS) features are designed to enable maximum system uptime for mission-critical applications by helping to safeguard servers against certain types of memory errors. The Dell PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers provide enhanced memory RAS features such as chip fail correct, spare-bank memory, and memory mirroring. This article describes the memory subsystems on these PowerEdge servers, providing technical details about their RAS features and explaining how administrators can enable RAS features and monitor their servers for memory failures.

Understanding the memory subsystem on Dell PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers

Dell PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers use the same type of memory subsystem, which includes the Intel E7520 Memory Controller Hub (MCH), two memory channels (A and B), and three dual in-line memory module (DIMM) slots per channel. The six DIMM slots form three logical banks (banks 1, 2, and 3), and the memory bus technology is two-way interleaved double data rate 2 memory at 400 MHz (DDR2-400). Figure 1 shows the architecture of this memory subsystem. The memory subsystem also implements error-correcting code (ECC), which is discussed in greater detail in the following section.

ECC memory

ECC-capable memory can detect and correct single-bit data errors and can detect double-bit data errors—a capability known as Single Error Correct/Double Error Detect (SEC/DED). ECC uses additional bits called ECC bits to calculate this parity information, which is stored along with the data bits in system memory. The MCH is responsible for calculating this information. When data is read from memory during a memory read operation, the MCH

 Reprinted from *Dell Power Solutions,* August 2005. Copyright © 2005 Dell Inc. All rights reserved.

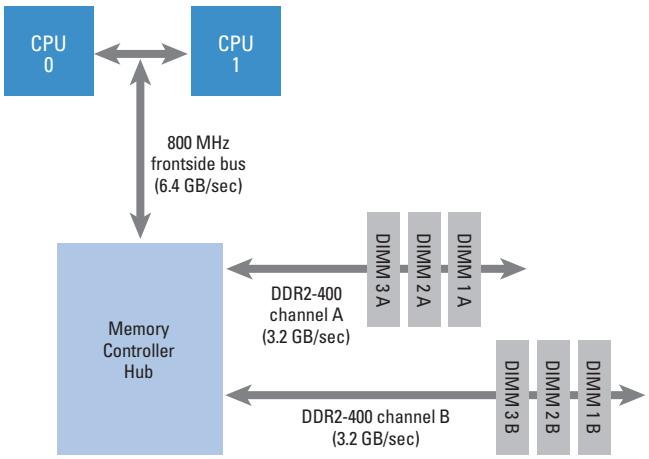

Figure 1. Memory subsystem on PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers

computes the new parity and compares it with the stored parity. If there is a mismatch, then a mechanism within the MCH detects the faulty bit. The faulty bit is "flipped" from "1" to "0" or from "0" to "1" and returned to the host. The error is then reported in the systems management logs. If a multibit error occurs, it is detected and logged but not corrected.

The memory subsystem on PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers implements standard ECC. The memory bus has 64 data lines and eight ECC lines.

Minimizing memory errors with RAS features

Memory errors are characterized as either soft or hard. *Soft errors* are transient and occasional; *hard errors* are permanent and are found in the silicon or in metallization of the dynamic RAM (DRAM) packaging. Regardless of the type of error, the number of data bits in error determines the system behavior. Single-bit errors are usually correctable in an ECC memory system. Multibit errors may be fatal if a system cannot recover from them.

The following sections describe the RAS features available in PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers and how they can help minimize memory errors. Chip fail correct helps overcome one class of multibit errors. Spare-bank memory and memory mirroring are other advanced mechanisms that enable systems to handle memory errors.

Chip fail correct

A DIMM has multiple identical DRAM devices. Each DRAM device on a DIMM can have 4 or 8 data lines. The corresponding DIMM is called a x4 or x8 DIMM, respectively. In a x4 ECC DIMM, 18 DRAM devices are needed to form the width of the memory bus (18 × 4 = 72 bits [64 data bits and 8 ECC bits]). These 18 DRAM devices form a *rank*. Similarly, for a x8 single-rank DIMM, nine DRAM devices are needed. The PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers support single-rank as well as dual-rank DIMMs. A maximum of four ranks are supported per memory channel.

Chip fail correct enables a system to withstand a multibit error within a DRAM device on a DIMM. On the PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers, chip fail correct is fully supported on x4 DIMMs and partially supported on x8 DIMMs. On x4 DIMMs, chip fail correct can rectify the error when all the bits of a DRAM are in error; on x8 DIMMs, this feature works only when up to four data lines are in error. In chip fail correct, every bit of a DRAM is part of a separate *ECC word*. An ECC word consists of data bits and ECC bits. The ECC bits are computed for each ECC word and stored. If multiple bits of a DRAM are in error, then all the bits can be corrected by the ECC algorithm because the chip fail correct feature recognizes them as a single-bit error for each ECC word.

The PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers include memory RAS capabilities that are designed to help minimize system downtime caused by memory failures.

On the PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers, in two-channel interleaved mode, the 144 bits are divided into four 36-bit ECC words, as shown in Figure 2. The BIOS enables the chip fail correct feature in the MCH when it detects x4 DIMMs installed.

Spare-bank memory

The spare-bank memory feature is implemented in the MCH and the system BIOS, enabling administrators to configure a spare row, or

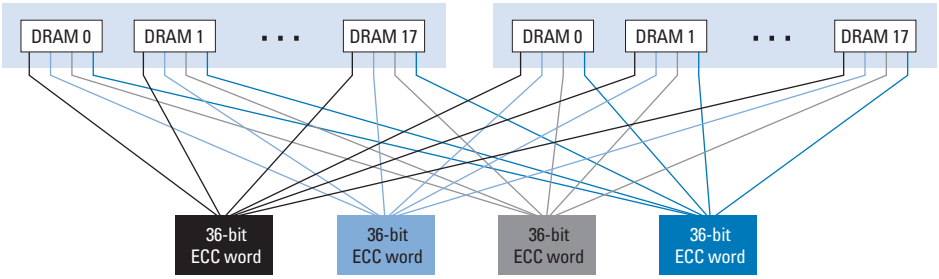

Figure 2. Bits divided into ECC words in two-channel interleaved mode

 Reprinted from *Dell Power Solutions,* August 2005. Copyright © 2005 Dell Inc. All rights reserved.

bank, of memory. This feature is available on the PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers only when all three banks of memory are populated with identical DIMMs. Administrators can configure this feature in the system BIOS setup (by pressing F2 to access the BIOS settings). The BIOS detects the memory configuration during the power-on self-test (POST) and provides the administrator with the option to enable this feature.

When this feature is enabled, only banks 1 and 2 are available to the administrator—as the spare, bank 3 is not in use. If the number of correctable errors that occur within a certain time frame on a particular DIMM in either bank 1 or bank 2 exceeds a threshold set in the BIOS, the BIOS will instruct the MCH to copy all data from the failing bank to bank 3 and to remap the memory in the system. After the copy process has completed, the failing bank is no longer used. The system uses bank 3 instead.

The impact of this feature is that one-third of the installed memory is not available to the system at any one time. In return, the system can benefit from an intelligent memory fault-tolerance mechanism that enables recovery from a large number of continuous correctable memory errors.

Memory mirroring

Like spare-bank memory, memory mirroring is implemented in the MCH and the system BIOS. Administrators can configure the system to maintain a mirrored copy of the data in memory, similar to RAID-1 in storage technology. On the PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers, this feature is available only when banks 1 and 2 are populated with identical DIMMs—and bank 3 must be empty. This feature is configurable in the system BIOS setup.

When an administrator enables this feature, banks 1 and 2 are configured as mirrored copies of each other. One bank is the primary copy while the other bank is the secondary copy. When data is written to memory, it is written to both banks. However, on a memory read, data is read back from the primary copy. If an uncorrectable error is detected on the read operation, the BIOS switches the primary and secondary banks. The system then reads from the new primary bank. A lost memory redundancy event is logged in the system event log (SEL).

The impact of this feature is that half of the installed memory is not available to the system at any one time. In return, the system can benefit from protection against uncorrectable memory errors that might otherwise result in unexpected system downtime and possibly data loss.

Enabling RAS features in Dell PowerEdge servers

Although chip fail correct is automatically enabled by the BIOS in the MCH, spare-bank memory and memory mirroring features must be configured by the administrator. Spare-bank memory and memory mirroring features are mutually exclusive; administrators can enable one or the other, but not both at the same time. Furthermore, these two features are available to administrators in the BIOS setup only if the supported memory configurations are installed.

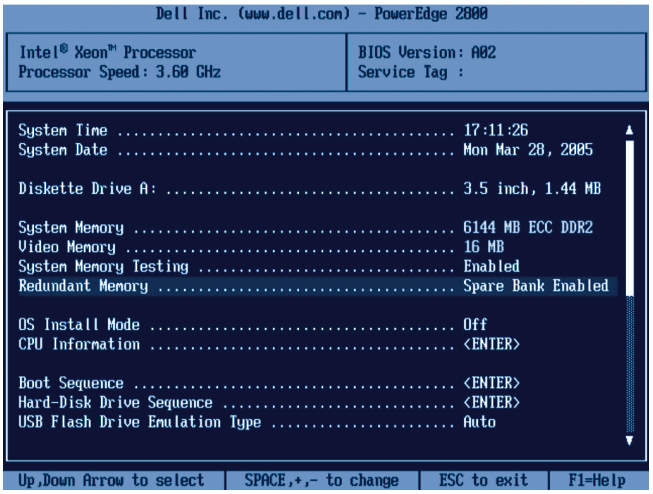

Figure 3. BIOS setup screen showing that spare-bank memory is enabled

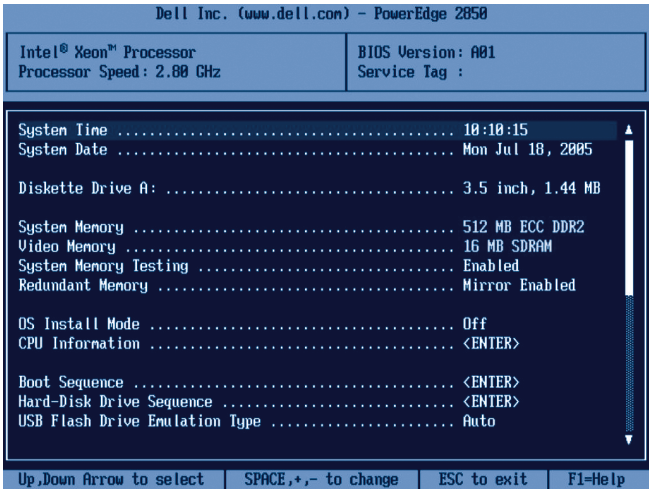

Figure 4. BIOS setup screen showing that memory mirroring is enabled

When configuring spare-bank memory and memory mirroring features, administrators should press F2 to enter the system BIOS setup. The memory setting is listed under "Redundant Memory" on the BIOS setup screen, as shown in Figures 3 and 4.

Monitoring memory events

The BIOS logs memory error, spare-bank memory, and mirroring events in the SEL in the baseboard management controller (BMC). Administrators can view the SEL through Dell OpenManage™ Server Administrator or the Web-based graphical user interface of the Dell Remote Access Controller (DRAC). Figure 5 shows the

 Reprinted from *Dell Power Solutions,* August 2005. Copyright © 2005 Dell Inc. All rights reserved.

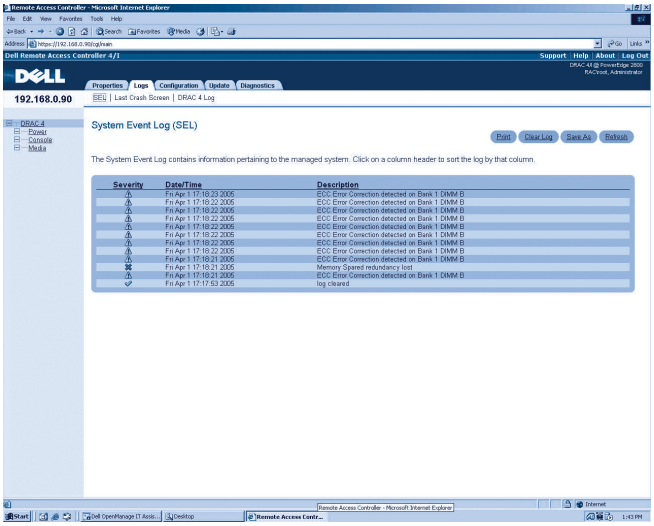

Figure 5. System event log showing memory events, accessed through the DRAC

SEL when accessed through the DRAC. Administrators can then perform a scheduled shutdown, replace faulty DIMMs, and get systems up and running.

Enhancing memory reliability and system availability

The PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850 servers include memory RAS capabilities that are designed to help minimize system downtime caused by memory failures. Administrators can easily enable these features and then monitor memory failure events by viewing system logs. By equipping PowerEdge servers with fault-tolerant memory capabilities, Dell helps data center administrators enhance reliability and availability of business-critical systems.

References

Locklear, David. "Chipkill Correct Memory Architecture." Dell Enterprise Systems Group technology brief, August 2000. www.ece.umd.edu/courses/enee759h.S2003/references/chipkill.pdf.

Qingsong Li is a product development engineer in the Dell Enterprise Server Product Group. He has a B.S. in Electrical Engineering from Harbin Institute of Technology in China and an M.S. in Electrical Engineering from The Ohio State University.

Utpal Patel is a test engineer in the Server Product Test Group at Dell. He has a bachelor's degree in Computer Engineering from Mumbai University in India and an M.S. in Computer Networking from North Carolina State University.

FOR MORE INFORMATION

Dell PowerEdge 1850, PowerEdge 2800, and PowerEdge 2850:
www.dell.com/poweredge

 Reprinted from *Dell Power Solutions,* August 2005. Copyright © 2005 Dell Inc. All rights reserved.